

Gewichtung im SOEP

Workshop zur Nutzung des SOEP
März 2010

Dr. Martin Kroh

Deutsches Institut für Wirtschaftsforschung (DIW Berlin)
Sozio-Oekonomische Panel Studie (SOEP)
Mohrenstr. 58, 10117 Berlin
Tel. 030 89789 678, Email. mkroh@diw.de

- 1 Grundlagen der Gewichtung
 - 1 Selektionsproblem
 - 2 Stichprobenziehung und Designgewichtung
 - 3 (Panel-)Non-Response und Non-Response Gewichtung
- 2 Gewichtung im SOEP
 - 1 Die Stichprobenentwicklung des SOEP
 - 2 Design-Gewichte
 - 3 Erste-Welle-Gewichte
 - 4 Gewichte der Welle 2+
- 3 Umgang mit Gewichten
 - 1 Quer- und Längsschnittgewichtung
 - 2 Varianzschätzung
 - 3 Pro & Contra Gewichtung

Stichprobenselektivität

- Empirische Sozialforschung
 - Ziel sind Aussagen über größere Grundgesamtheit.
 - Inferenz auf Basis kleiner Untersuchungsgruppe.
 - Analyseverfahren unterstellen oft einfache Zufallsstichproben.
- Aber ...
 - Alternative Stichprobendesigns haben Vorteile (selection by design)
 - Gezogene Einheiten verweigern die Teilnahme an Untersuchung (self-selection)
 - Ignorieren von Sample Selection *kann* Parameterschätzungen verzerren.
- Arten von Lösungen
 - Complete case analysis (keine Kompensation)
 - Modell-basierte Lösungen (z.B. Heckman Selektionsmodell)
 - Imputation (Ersetzen fehlender Daten)
 - Gewichtung

Selektions-Mechanismen (Little & Rubin 2002)

Stichprobenselektivität als Problem fehlender Daten.

- Missing completely at random (MCAR)
 - Fehlende Daten (=Selektion) hängen weder von beobachteten Daten ab, noch den unbeobachteten Werten der interessierenden Variable.
 - Zufallsstichprobe, Zufallsexperiment
 - Beispiel: Untersuchung von Augenfarbe und geringere Responserate von jungen Personen.
- Missing at random (MAR)
 - Fehlende Daten (=Selektion) hängen von beobachteten Daten ab, aber nicht den unbeobachteten Werten der interessierenden Variable.
 - Beispiel: Untersuchung von Einkommen und geringere Responserate von jungen Personen.
- Not missing at random (NMAR)
 - Fehlende Daten (=Selektion) hängen von unbeobachteten Werten der interessierenden Variable ab.
 - Beispiel: Untersuchung von Einkommen und geringere Responserate von Beziehern geringer Einkommen.

Anwendbarkeit unterschiedlicher Lösungen

- Complete case analysis
 - MCAR
- Model-basierte Lösungen
 - NMAR, aber Annahmen bzgl. der Fehlerverteilung und/oder exogener Beobachtungen (Heckman Selektionsmodell, Tobit, etc.).
 - Häufig bei Item Non-Response verwendet. Spezifische Lösung für *eine* abhängige Variable.
- Imputation (von Variablen)
 - MAR, aber ...
 - Häufig bei heterogenem Muster fehlender Daten verwendet. Allgemeine Lösung für z.B. Item-Non-Response. Imputieren *aller* Angaben (Unit Non-Response) extrem komplex.
- Gewichtung (von Einheiten)
 - MAR, aber ...
 - Häufig bei monotonem Non-Response Muster verwendet. Allgemeine Lösung z.B. für Unit Non-Response in Panel-Daten.

Beispiel: Arbeitslosigkeitserfahrung

Simulierte Daten!

- Grundgesamtheit
 - Kohorte 1980 zum Alter 28 in Deutschland (~ 1 Mio.)
 - Wieder/Befragung 2008 und 2009
- Stichprobe
 - 1000 Befragte
 - Schichtung von Personen mit/ohne Migrationshintergrund in Ost- und Westdeutschland um getrennte Analysen zu gewährleisten.
- Erkenntnisinteresse
 - Arbeitslosigkeitserfahrung und Lebenszufriedenheit.

Beispiel: Grundgesamtheit

- Kohorte 1984 in 2009

	West	Ost	Gesamt
Kein Migrationsh.	640.000	185.000	825.000
Migrationsh.	160.000	15.000	175.000
Gesamt	800.000	200.000	1.000.000

- Anteile in Grundgesamtheit

- Ostdeutsche: $\frac{200.000}{1.000.000} = 20\%$
- Migranten: $\frac{175.000}{1.000.000} = 17.5\%$
- Migranten unter Ostdeutschen: $\frac{15.000}{200.000} = 7.5\%$
- Migranten unter Westdeutschen: $\frac{160.000}{800.000} = 20\%$

Beispiel: Proportional geschichtete Stichprobenziehung

- Ziehungswahrscheinlichkeit von $\frac{1}{1.000}$

	West	Ost	Gesamt
Kein Migrationsh.	640	185	825
Migrationsh.	160	15	175
Gesamt	800	200	1.000

- Anteile in der Stichprobe
 - Ostdeutsche: $\frac{200}{1.000} = 20\%$
 - Migranten: $\frac{175}{1.000} = 17.5\%$
 - Migranten unter Ostdeutschen: $\frac{15}{200} = 7.5\%$
 - Migranten unter Westdeutschen: $\frac{160}{800} = 20\%$

Beispiel: Disproportional geschichtete Stichprobenziehung

- Schichtung

	West	Ost	Gesamt
Kein Migrationsh.	250	250	500
Migrationsh.	250	250	500
Gesamt	500	500	1.000

- Ziehungswahrscheinlichkeit in Schichten

- Gesamt: $\frac{1.000}{1.000.000} = \frac{1}{1.000}$
- Nicht-Migranten in West-D.: $\frac{250}{640.000} = \frac{1}{2.560}$
- Migranten in West-D.: $\frac{250}{160.000} = \frac{1}{640}$
- Nicht-Migranten in Ost-D.: $\frac{250}{185.000} = \frac{1}{740}$
- Migranten in Ost-D.: $\frac{250}{15.000} = \frac{1}{60}$

Beispiel: Gewichte = Inverse Ziehungs'wkt

- Einheiten, die aus der Grundgesamtheit mit der Wahrscheinlichkeit π_j gezogen werden, "repräsentieren" π_j^{-1} Einheiten der Grundgesamtheit und sollten somit ein Gewicht von π_j^{-1} in der Schätzung von Populations-Parametern haben.
- Ziehungswahrscheinlichkeit in den j Schichten
 - Nicht-Migranten in West-D.: $\pi_{i1}^{-1} = 2.560$
 - Migranten in West-D.: $\pi_{i2}^{-1} = 640$
 - Nicht-Migranten in Ost-D.: $\pi_{i3}^{-1} = 740$
 - Migranten in Ost-D.: $\pi_{i4}^{-1} = 60$

Beispiel: Verteilung der Gewichte

Stratum	N(samplewght)	mean(samplewght)	sum(samplewght)
West, Native	250	2,560	640,000
West, Migrant	250	640	160,000
East, Native	250	740	185,000
East, Migrant	250	60	15,000
Total	1,000	1,000	1,000,000

Beispiel: Arbeitslosigkeitserfahrung nach Ziehungsschicht (=Strata)

=> ungewichtet

Stratum	Mean(A'losigkeit)
West, Native	0.10
West, Migrant	0.27
East, Native	0.22
East, Migrant	0.37
Total	0.24

=> gewichtet mit Designgewichten

Stratum	Mean(A'losigkeit)
West, Native	0.10
West, Migrant	0.27
East, Native	0.22
East, Migrant	0.37
Total	0.15

Ziehungs'wkt und Designgewichtung

- Selektionsindikator c_i mit

$$c_i = \begin{cases} 1 & \text{Einheit } i \text{ in Stichprobe gezogen,} \\ 0 & \text{sonst} \end{cases}$$

- Einfache Zufallsstichprobe

$$f(c) = \begin{cases} \binom{N}{n}^{-1} & \text{if } \sum_i c_i = n \\ 0 & \text{sonst} \end{cases}$$

- Geschichtete Zufallsstichprobe (z nicht-überlappend und erschöpfend)

$$f(c|z) = \begin{cases} \prod_{j=1}^J \binom{N_j}{n_j}^{-1} & \text{falls } \sum_{i:z_i=j} c_i = n_j \text{ für alle } j \\ 0 & \text{sonst} \end{cases}$$

Inverse Selection Probability Weighting Scheme

- Zu schätzender Kennwert z.B. Merkmalssumme von Y :

$$t_y = \sum_{i=1}^N y_i$$

- “Inverse Selection Probability” Gewichtung

$$w_i = \frac{c_i}{\pi_i} \quad \text{mit} \quad \pi_i = E(c_i) = \Pr(c_i = 1).$$

- Unverzerrte Schätzung durch Gewichtung (Horvitz-Thompson, 1952)

$$\hat{t}_y = \sum_{i=1}^N w_i y_i \quad \text{mit} \quad E(w_i) = 1.$$

Bekannte und unbekannte Ziehungs'wkten

- Im Fall der geschichteten Stichprobenziehung sind die Ziehungs'wkten und somit die Designgewichte π_i^{-1} durch den Forscher vorgegeben und somit bekannt.
- In vielen Fällen hängt die Beobachtung jedoch von der freiwilligen Teilnahme der Untersuchungseinheiten ab. Beobachtungs'wkten und somit Non-Response Gewichte $\hat{\phi}_i^{-1}$ müssen geschätzt werden.
- Schätzung der Beobachtungs'wkten erfordert Erhebung von Informationen über Teilnehmer *und* Nicht-Teilnehmer.

Erweiterung der “Inverse Selection Probability” Gewichtung

- Einfachster Ansatz: Gewicht ist die inverse Beobachtungsrate in Schichten des Ziehungsdesigns, $\hat{\phi}_j^{-1}$.
- Non-Response in Querschnitten oft problematisch, da über Schichtungsmerkmale hinaus wenige Informationen über Nicht-Teilnehmer zugänglich sind.
 - Aggregatinformationen zu PSU's (Region, Schule, etc.)
 - Proxy-Informationen der Interviewer, z.B. zum Wohnumfeld
 - Screening-Information von früherer Stufe der Stichprobenziehung
 - Kurzfragebogen für Verweigerer.
 - Externe Informationen zu Verteilungen in Grundgesamtheit (Poststratifizierung, Raking)
- Ausfälle in Panel-Studien aufgrund der Vielzahl an Informationen zu Ausfällen und Teilnehmern aus vorherigen Wellen viel besser zur Schätzung von Ausfallgewichten geeignet.

Beispiel: Panelausfälle

- Non/Response in Welle 2
 - Ein Drittel der Befragten aus 2008 können in 2009 nicht wiederbefragt werden.
- Beobachtungs'wkt in Welle 2 korreliert (MAR) mit ...
 - Schichtungsmerkmalen
 - Arbeitslosigkeitserfahrung (-)
- Vorgehen
 - Schätzung eines Modells (logit/probit) der Beobachtung in Welle 2 auf Basis aller Informationen der Welle 1.
 - Die inverse vorhergesagte Response'wkt ist das geschätzte Gewicht der Welle 2, $\hat{\phi}_i^{-1}$.
 - Das Produkt aus Designgewicht und Non-Response-Gewicht korrigiert für Schichtung und Ausfall in Welle 2, $(\pi_i \times \hat{\phi}_i)^{-1}$.

Beispiel: Response-Rate und Response-Modell

Stratum	A'los 2008		
	No	Yes	Total
West, Native	0.83	0.56	0.80
West, Migrant	0.48	0.18	0.40
East, Native	0.90	0.59	0.83
East, Migrant	0.80	0.37	0.64
Total	0.76	0.38	0.67

logit regression

Number of obs = 1000
Pseudo R2 = 0.18

Response 2009	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
A'los 2008	-1.710772	.1763381	-9.70	0.000	-2.056388 -1.365156
West, Migrant	-1.670855	.2131702	-7.84	0.000	-2.088661 -1.253049
East, Native	.5295863	.2466893	2.15	0.032	.0460841 1.013088
East, Migrant	-.3357836	.223581	-1.50	0.133	-.7739943 .1024271
_cons	1.626398	.1667497	9.75	0.000	1.299575 1.953222

Beispiel: Kennwerte der Non-Response-Gewichte

Stratum		A'los 2008		
		No	Yes	Total
West, Native	n	186	14	200
	mean	1.20	2.09	1.26
	sum	223	29	252
West, Migrant	n	87	12	99
	mean	2.05	6.78	2.62
	sum	178	81	259
East, Native	n	175	33	208
	mean	1.12	1.64	1.20
	sum	195	54	249
East, Migrant	n	126	34	160
	mean	1.28	2.52	1.54
	sum	161	86	246
Total	n	574	93	667
	mean	1.32	2.69	1.51
	sum	756	251	1007

Beispiel: Produkt der Design- und Non-Response-Gewichte

Stratum	mean(samplew.)	mean(responsew.)	mean(jointw.)
West, Native	2560	1.26	3223
West, Migrant	640	2.62	1677
East, Native	740	1.20	887
East, Migrant	60	1.54	92
Total	1000	1.51	1514
	(n=1000)	(n=667)	(n=667)

Beispiel: Kennwerte der Beobachtungsgewichte (Produkt)

Stratum		A'los 2008		
		No	Yes	Total
West, Native	n	186	14	200
	mean	3063	5345	3223
	sum	569790	74835	644626
West, Migrant	n	87	12	99
	mean	1309	4342	1677
	sum	113891	52107	165998
East, Native	n	175	33	208
	mean	826	1214	887
	sum	144495	40066	184560
East, Migrant	n	126	34	160
	mean	77	151	92
	sum	9640	5145	14785
Total	n	574	93	667
	mean	1460	1851	1514
	sum	837816	172153	1009969

Erweiterung der “Inverse Selection Probability” Gewichtung

- Indikator für die Selektion per Design s_i mit

$$s_i = \begin{cases} 1 & \text{Einheit } i \text{ in Stichprobe gezogen,} \\ 0 & \text{sonst} \end{cases}$$

- Indikator für die Selbstselektion r_i with

$$r_i = \begin{cases} 1 & \text{Einheit } i \text{ nimmt in Studie teil,} \\ 0 & \text{sonst} \end{cases}$$

- Beobachtungs'wkt der Einheit i :

$$\Pr(c_i = 1) = \Pr(s_i = 1) \times \Pr(r_i = 1 | s_i = 1) = [\pi_i \times \hat{\phi}_i]$$

wobei

- $\Pr(s_i = 1)$ bekannt ist,
- $\Pr(r_i = 1 | s_i = 1)$ unbekannt ist und geschätzt werden muss und
- $1 / \{\Pr(s_i = 1) \times \hat{\Pr}(r_i = 1 | s_i = 1)\}$ das gemeinsame Gewicht ist.

Poststratifizierung und Raking

- Idee
 - (Marginale) Verteilungen der Grundgesamtheit durch externe Quelle bekannt (oft Mikrozensus).
 - Gewichte werden so angepasst dass gewichtete Verteilungen in beobachteter Stichprobe den "wahren" Verteilungen der Grundgesamtheit entsprechen.
- Methoden
 - Poststratifizierung: Wenige Merkmale, alle Zellen bekannt.
 - Raking: Mehr Merkmale mit bekannte Randverteilungen und unbekanntes Zellenbesetzungen (Iterative Proportionale Anpassungs-Prozedur).
- Situationen in denen Poststrat./Raking attraktiv ist.
 - Kaum Informationen zu Nicht-Teilnehmern.
 - Sozialpolitische Evaluationen: möglichst gute Approximation bestimmter Merkmale der Grundgesamtheit.
 - Häufige Merkmale: Alter, Geschlecht, Region

Stichproben des SOEP

- ① Stichprobe A (Westdeutsche)
 - Erste Welle: 1984
 - Grundgesamtheit: Privat-HH mit deutschem HH-Vorstand in Westdeutschland und West-Berlin
 - Befragte HH: 4528
- ② Stichprobe B (Ausländer)
 - Erste Welle: 1984
 - Grundgesamtheit: Privat-HH mit türkischem, italienischem griechischem, jugoslawischem und spanischem HH-Vorstand in Westdeutschland und West-Berlin
 - Befragte HH: 1393
- ③ Stichprobe C (Ostdeutsche)
 - Erste Welle: 1990
 - Grundgesamtheit: Privat-HH in DDR und Ost-Berlin
 - Befragte HH: 2179

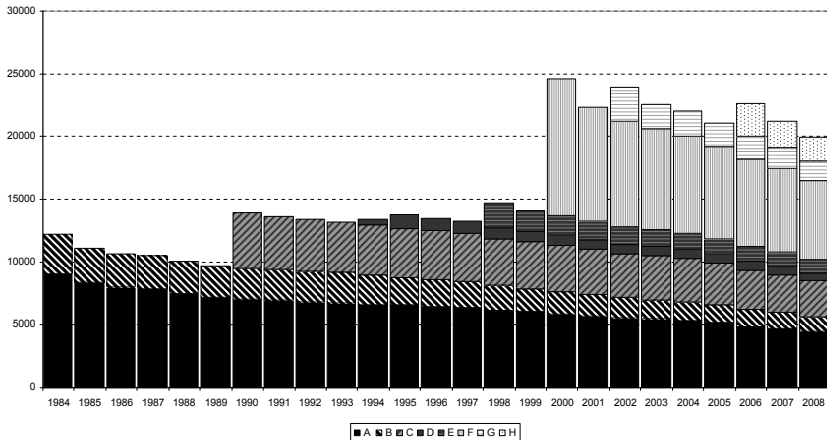
Stichproben des SOEP

- ④ Stichprobe D (Zuwanderer)
 - Erste Welle: 1994/5
 - Grundgesamtheit: Privat-HH, die seit 1984 nach (West-)Deutschland kamen
 - Befragte HH: 522
- ⑤ Stichprobe E (Ergänzung I)
 - Erste Welle: 1998
 - Grundgesamtheit: Privat-HH in Deutschland
 - Befragte HH: 1056
- ⑥ Stichprobe F (Ergänzung II)
 - Erste Welle: 2000
 - Grundgesamtheit: Privat-HH in Deutschland
 - Befragte HH: 6052

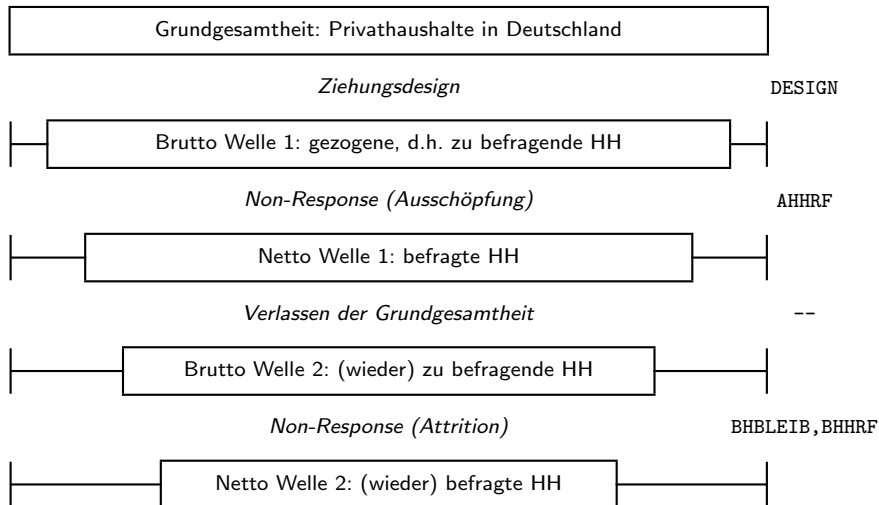
Stichproben des SOEP

- ⑦ Stichprobe G (Hocheinkommen)
 - Erste Welle: 2002
 - Grundgesamtheit: Privat-HH mit einem monatl. Netto-Einkommen von mehr als 7500 DM
 - Befragte HH: 1124
- ⑧ Stichprobe H (Ergänzung III)
 - Erste Welle: 2006
 - Grundgesamtheit: Privat-HH in Deutschland
 - Befragte HH: 1506

Stichprobengröße SOEP 1984–2008, Personen



Stichprobenentwicklung



Grundgesamtheit → Brutto 1: Stichprobendesign (DESIGN)

- Klumpung (PSU)
 - Primäre Ziehungseinheiten (primary sampling unit) Regionen (Wahlkreise)
 - Ausnahme D und G: bundesweites Screening
 - Je etwa 250 (E) bis 984 (F) primäre Ziehungseinheiten.
 - Sekundäre Ziehungseinheiten (secondary sampling unit) Haushalte
 - Ausnahmen B: Personen aus Ausländerregister
 - Je etwa 2029 (E) bis 11871 (F) sekundäre Ziehungseinheiten.
- Für Gewichtung weinger relevant, jedoch für Varianzschätzung wichtig.

Grundgesamtheit → Brutto 1: Stichprobendesign (DESIGN)

- Schichtung (STRAT)
 - Keine Stratifizierung in A, C, E und H
 - Stratifizierung nach Zuwanderer(-gruppen) in B, D und F
 - Stratifizierung nach Einkommen und Region (Ost/West) in G
 - Integration von Sub-Stichproben in gesamtes SOEP Form von Stratifizierung
- Ziehungs'wkten ($\text{DESIGN} = 1/\text{Pr}(s_i = 1)$)
 - Hohe Ziehungs'wkt (=geringe Gewichte) NBL (C und G)
 - Hohe Ziehungs'wkt (=geringe Gewichte) Hocheinkommen (G)
 - Hohe Ziehungs'wkt (=geringe Gewichte) Zuwanderer (B, D und F)

Designgewichte im Datensatz DESIGN nach Stichproben

Stichprobenart	Min	Mean	Max
A Deutsche (West)	3344	3344	3344
B Auslaender (West)	68	546	1697
C Deutsche (Ost)	1900	1900	1900
D Zuwanderer 1984-93	2946	3279	3693
E Ergaenzung 1998	19081	19081	19081
F Ergaenzung 2000	1980	3406	3519
G Hohe Einkommen 2002	490	754	956
H Ergaenzung 2006	10432	10432	10432
Total	68	4296	19081

Brutto 1 → Netto 1: Non-Response, Ausschöpfung (HHRF)

- Wie entsteht der HRF in Welle 1 (z.B. $AHHRF = 1 / \{\Pr(s_i = 1) \times \hat{Pr}(r_i = 1 | s_i = 1)\}$)?
 - 1 Ausfallanalyse auf Basis von Informationen zu Brutto 1
 - Interviewerangaben zum Wohnumfeld und Haushalt
 - Befragungsinformationen aus Screening
 - Regionalinformationen zu Kreisen
 - Regionalinformationen zur Nachbarschaft (MICROM)
 - 2 Abweichungen von bekannten Verteilungen der Grundgesamtheit (Post-Stratifizierung)
 - Mikrozensus-Informationen zu Altersstruktur, Geschlechter-Verteilung, Nationalitäten, etc. als "wahre" Werte der Grundgesamtheit angenommen.
 - Summe der Gewichte werden den bekannten Summen in Grundgesamtheit angepasst.

Brutto 1 → Netto 1: Non-Response, Ausschöpfung (HHRF)

- Wo sind die HRF's der Welle 1 abgelegt?
 - AHHRF für A und B (1984)
 - GHHRF für C (1990)
 - LHHRF für D (1995)
 - OHHRFE für E (1998)
 - QHHRFF für F (2000)
 - SHHRFG für G (2002)
 - WHHRFH für H (2006) im Datensatz HHRF ...
 - ... entsprechendes gilt für die Personen-HRF's in PHRF.
- Beachte: Gewichte, die die Stichprobeneinheiten auf Summen der Grundgesamtheit "hochrechnen" werden im SOEP Hochrechnungsfaktoren ($\$HHRF$ und $\$PHRF$) genannt, können jedoch leicht durch eine lineare Transformation auf die Summe der Stichprobengröße reduziert werden ohne die relative Gewichtung der Einheiten zueinander zu verändern.

Beispiel: Bildungsabschlüsse 1984 ungewichtet

```
. tab casmin84
```

CASMIN-Klassifikation	Freq.	Percent	Cum.
(1a) inadequately completed	1,067	8.88	8.88
(1b) general elementary school	2,975	24.76	33.64
(1c) basic vocational qualification	4,319	35.94	69.58
(2b) intermediate general qualificatio	591	4.92	74.49
(2a) intermediate vocational	1,580	13.15	87.64
(2c_gen) general maturity certificate	322	2.68	90.32
(2c_voc) vocational maturity certifica	370	3.08	93.40
(3a) lower tertiary education	252	2.10	95.50
(3b) higher tertiary education	541	4.50	100.00
Total	12,017	100.00	

Beispiel: Bildungsabschlüsse 1984 gewichtet

```
. tab casmin84 [aweight=aphrf]
```

CASMIN-Klassifikation	Freq.	Percent	Cum.
(1a) inadequately completed	300.830687	2.50	2.50
(1b) general elementary school	2,943.2102	24.49	27.00
(1c) basic vocational qualification	4,745.196	39.49	66.48
(2b) intermediate general qualificatio	566.33219	4.71	71.20
(2a) intermediate vocational	1,782.657	14.83	86.03
(2c_gen) general maturity certificate	353.460538	2.94	88.97
(2c_voc) vocational maturity certifica	447.426998	3.72	92.69
(3a) lower tertiary education	301.084046	2.51	95.20
(3b) higher tertiary education	576.80228	4.80	100.00
Total	12,017	100.00	

Beispiel: Bildungsabschlüsse 1984 gewichtet

```
. tab casmin84 [fweight=round(aphrf)]
```

CASMIN-Klassifikation	Freq.	Percent	Cum.
(1a) inadequately completed	1,235,350	2.50	2.50
(1b) general elementary school	12,086,105	24.49	27.00
(1c) basic vocational qualification	19,485,818	39.49	66.48
(2b) intermediate general qualificatio	2,325,593	4.71	71.20
(2a) intermediate vocational	7,320,331	14.83	86.03
(2c_gen) general maturity certificate	1,451,471	2.94	88.97
(2c_voc) vocational maturity certifica	1,837,338	3.72	92.69
(3a) lower tertiary education	1,236,389	2.51	95.20
(3b) higher tertiary education	2,368,604	4.80	100.00
Total	49,346,999	100.00	

Beispiel: Bildungsabschlüsse 1984 gewichtet

```
. tab casmin84 [iweight=aphrf]
```

CASMIN-Klassifikation	Freq.	Percent	Cum.
(1a) inadequately completed	1,235,338	2.50	2.50
(1b) general elementary school	12086069.5	24.49	27.00
(1c) basic vocational qualification	19485787.4	39.49	66.48
(2b) intermediate general qualificatio	2,325,600	4.71	71.20
(2a) intermediate vocational	7,320,346	14.83	86.03
(2c_gen) general maturity certificate	1,451,459	2.94	88.97
(2c_voc) vocational maturity certifica	1,837,325	3.72	92.69
(3a) lower tertiary education	1,236,379	2.51	95.20
(3b) higher tertiary education	12,368,594.8	4.80	100.00
Total	49346898.8	100.00	

Beispiel: Bildungsabschlüsse 1984 gewichtet

```
. tab casmin84 [iweight=aphrfneu]
```

CASMIN-Klassifikation	Freq.	Percent	Cum.
(1a) inadequately completed	300.830679	2.50	2.50
(1b) general elementary school	2,943.2101	24.49	27.00
(1c) basic vocational qualification	4,745.1959	39.49	66.48
(2b) intermediate general qualificatio	566.332176	4.71	71.20
(2a) intermediate vocational	1,782.657	14.83	86.03
(2c_gen) general maturity certificate	353.460528	2.94	88.97
(2c_voc) vocational maturity certifica	447.426987	3.72	92.69
(3a) lower tertiary education	301.0840385	2.51	95.20
(3b) higher tertiary education	576.802264	4.80	100.00
Total	12,016.9997	100.00	

```
egen count=count(aphrf)  
egen sum=sum(aphrf)  
gen aphrfneu=aphrf*count/sum
```

Netto 1 → Brutto 2: Verlassen der Grundgesamtheit

- Wer soll in der zweiten Welle wieder befragt werden?
 - Haushalte der Welle 1, die weiterhin zur Grundgesamtheit gehören, also in Deutschland gemeldet sind.
 - Abspaltungen von Haushalten der Welle 1, die zur Grundgesamtheit gehören, also in Deutschland gemeldet sind.
- Wer soll in der zweiten Welle nicht mehr befragt werden?
 - Haushalte der Welle 1, die nicht mehr zur Grundgesamtheit gehören, da sie sich z.B. durch Tod aufgelöst haben oder ins Ausland verzogen sind.
- Was ist mit Zugängen zur Grundgesamtheit zwischen Welle 1 und 2?
 - Notwendigkeit von Auffrischungen auf Haushaltesebene (vgl. E, F, H und insbes. D)
 - Zugänge in bestehende Haushalte, z.B. durch Geburt und Zuzug, erfassen teilweise Veränderungen der Grundgesamtheit auf Personenebene.

Brutto 2 → Netto 2: Non-Response, Attrition (\$HBLEIB)

- Wie entsteht der Bleibefaktor in Welle 2 (z.B. BHLEIB = $1 / \Pr(c_{it} = 1 | c_{i,t-1} = 1)$)?
 - 1 Ausfallanalyse auf Basis von Informationen zu Brutto 2 aus t-1
 - Merkmale des Interviews (Mode, Dauer, Interviewerwechsel, ...)
 - Merkmale des HH (Vermögen, HH-Größe, ...)
 - Aggregation von Personen-Merkmalen über HH (HH-Vorstand ist Ausländer, wenigstens eine Person im HH ist arbeitslos, ...)
 - Merkmale der Region (Arbeitslosigkeit auf Kreisebene, Umzugsvolumen in Straßenzug, ...)
 - 2 Kontakt'wkt und Response'wkt in zwei Schritten
 - Ein erstes Modell zur Bestimmung der Wkt, den HH wieder zu kontaktieren (Adressermittlung).
 - Ein zweites Modell zur Bestimmung der Wkt, dass ein HH zur Wiederbefragung bereit ist, gegeben der HH konnte wieder kontaktiert werden.
 - $(\text{Kontakt'wkt} \times \text{Response'wkt})^{-1} = \$BLEIB$

Bestimmung der Kontakt'wkt nach Stichproben in 2008

Table 3: Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2008.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Intercept	-4.85 (0.67) ***	-4.32 (0.46) ***	-4.47 (0.37) ***	-3.13 (0.79) ***	-4.69 (0.72) ***	-2.22 (0.31) ***	-4.97 (0.62) ***	-4.79 (0.46) ***
New HH	-2.81 (0.62) ***		-1.86 (0.37) ***			-2.18 (0.35) ***		
Moved HH		-1.41 (0.46) ***				-1.66 (0.37) ***	-1.77 (0.62) ***	-1.64 (0.46) ***
Temporary Drop-Out				-2.44 (0.79) ***		-1.09 (0.28) ***		
Single HH	-2.66 (0.80) ***							
Job-Change	-1.82 (0.60) ***							
Commuting					-1.80 (0.72) **			
Likelihood Ratio ($Pr > \chi^2$)	0.95	****	****	****	****	< 0.0001	****	****

Note. *** p < 0.01; ** p < 0.05; * p < 0.10; standard errors in parentheses. **** The specified and the saturated models are the same.

Bestimmung der Antwort'wkt nach Stichproben in 2008

Table 6a: Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2008.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Intercept	2.82 (0.28) ***	4.55 (0.77) ***	3.15 (0.34) ***	2.83 (0.77) ***	3.97 (0.51) ***	3.10 (0.18) ***	1.77 (0.49) ***	2.05 (0.50) ***
First Wave HH	0.19 (0.06) ***		0.30 (0.09) ***			0.21 (0.07) ***		0.84 (0.31) ***
New HH	-0.61 (0.19) ***							
HH Move						-0.21 (0.11) **		
Move Out						0.69 (0.32) **		
Additional Questionnaire 1		-0.50 (0.21) **						
Additional Questionnaire 2			0.39 (0.19) **					
Biography								0.49 (0.23) **
Face-to-Face	-2.04 (0.26) ***	-3.20 (0.73) ***	-2.26 (0.30) ***	-1.82 (0.74) **	-2.23 (0.48) ***	-1.87 (0.16) ***	-1.76 (0.38) ***	-1.87 (0.29) ***
CAPI		0.51 (0.20) **			0.43 (0.19) **			
Change in Interviewer	-0.28 (0.11) **				-0.84 (0.24) ***	-0.37 (0.11) ***		
Change in Interview-Mode						-0.37 (0.11) ***		
Non-Regular Interview	-2.30 (0.26) ***	-3.48 (0.72) ***	-2.55 (0.30) ***	-2.75 (0.73) ***	-2.82 (0.47) ***	-2.33 (0.14) ***	-2.39 (0.36) ***	-2.67 (0.21) ***
Phone Disclosed	0.36 (0.10) ***	0.43 (0.17) **	0.36 (0.12) ***				1.24 (0.29) ***	
Single HH								0.28 (0.12) **
4+ Person HH			-0.25 (0.11) **			-0.27 (0.07) ***		
Gender								-0.26 (0.10) ***
(Age 25-34)*(Old-HH)								-0.42 (0.16) ***
Unmarried			-0.31 (0.10) ***					
Divorce					-0.97 (0.30) ***			
(Separation)*(Old-HH)	-1.02 (0.38) ***					-0.94 (0.44) **		

Note. *** p < 0.01; ** p < 0.05; * p < 0.10; standard errors in parentheses.

Bestimmung der Antwort'wkt nach Stichproben in 2008

Table 6b: Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2008.

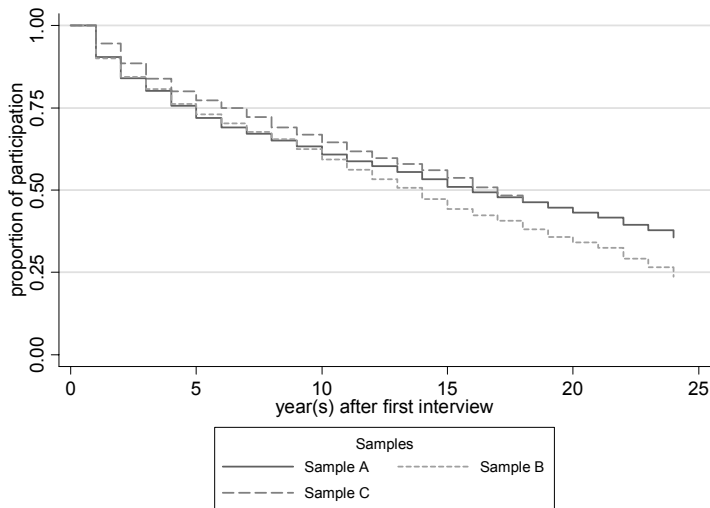
	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Employee					0.47 (0.19) **			
Self-Employed							-0.26 (0.13) **	
Irregular Employment					-0.61 (0.17) ***			
Job Worries	-0.27 (0.13) **		0.25 (0.09) ***					
Secondary Education			0.18 (0.09) **					
Tertiary Education					-0.47 (0.20) **			
Vocational Education				0.47 (0.21) **				
Voc. Ed. not Specified								-0.35 (0.15) **
Income 25							-0.27 (0.12) **	
Income 50						-0.15 (0.06) **		
Income 75			-0.25 (0.09) ***					
Income Not Specified			-0.60 (0.22) ***					-0.37 (0.10) ***
Savings					-0.51 (0.20) ***			
Wealth 50			-0.34 (0.13) ***					
Wealth not Specified	-0.36 (0.15) **		-0.25 (0.10) **					
Wealth Missing					-0.49 (0.17) ***			
Good Health						0.14 (0.07) **		
Disabled						-0.20 (0.07) ***		
Low Political Interest	-0.39 (0.18) **					-0.13 (0.05) **		

Note. *** p < 0.01; ** p < 0.05; * p < 0.10; standard errors in parentheses.

Brutto 2 → Netto 2: Non-Response, Attrition (\$HBLEIB)

- Wo sind die BLEIB's der Welle 2+ abgelegt ?
 - BHBLEIB bis YHBLEIB in HHRF
 - ... entsprechendes gilt für die Personen-BLEIB's in PHRF
- Im Gegensatz zu den HRF's rechnen die BLEIB's die Stichprobe nicht auf die Grundgesamtheit hoch, sondern stellen die Stichproben(-summen) des Samples in $t-1$ wieder her.
 - $\Pr(c_{it} = 1 | c_{i,t-1} = 1) \text{ idR } > 0.9$
 - $1 / \Pr(c_{it} = 1 | c_{i,t-1} = 1)$ somit nahe 1.

Verweildauer von Befragten der Welle 1 im SOEP



Brutto 2 → Netto 2: Non-Response, Attrition (\$HHRF)

- Wie entsteht der HRF in Welle 2?
 - 1 Bestimmung der "Rohgewichte"
 - $AHHRF \times BHBLEIB = BHHRF$
 - 2 Abweichungen von Verteilungen der Grundgesamtheit, Post-Stratifizierung (\$HHRF)
 - Haushaltsebene: Bundesland, Gemeindegröße, Haushaltsgröße, Eigentum (Mikrozensus).
 - Personenebene: Alter, Geschlecht, Nationalität (Mikrozensus).
 - Summe der Gewichte werden (iterativ) den Summen in Grundgesamtheit angepasst.
- Wo sind die HRF's der Welle 2 abgelegt?
 - BHHRF bis YHHRF in HHRF
 - ... entsprechendes gilt für die Personen-HRF's in PHRF

Verteilung der Querschnittsgewichte im SOEP, Haushalte

	ahhrf	bhhrf	chhrf	dhhrf	ehhrf	fhhrf	ghhrf	hhhrf	ihhrf	jhhrf	khhrf	lhhrf	mhhrf	nhhrf
p5	433.41	495.46	541.88	548.69	535.27	551.46	699.77	680.29	669.97	684.16	712.63	702.54	740.17	747.64
p10	600.48	681.65	757.63	795.13	804.65	822.26	1071.36	1048.48	1037.92	1054.62	1101.42	1112.20	1167.67	1204.37
p25	3775.97	3886.69	3615.83	3559.54	3574.10	3643.43	2216.64	2339.59	2331.49	2405.78	2410.23	2413.26	2408.34	2408.43
p50	4719.90	5081.43	5300.95	5405.57	5634.98	5855.85	4605.43	4697.75	4658.16	4690.17	4677.48	4376.16	4356.34	4331.58
p75	5657.32	6412.55	6825.70	7033.14	7544.82	7867.09	7042.49	7167.50	7141.61	7244.68	7279.06	6977.59	7004.35	7020.83
p90	7129.89	8460.74	9243.14	9583.02	10362.78	10851.10	9962.01	10265.52	10499.27	10755.37	11216.65	11069.13	11328.47	11886.85
p95	8305.07	10036.44	11123.83	11430.38	12537.17	13277.36	12363.62	12984.13	13642.81	14003.65	14604.14	14794.33	15257.40	15939.06
N	5921	5322	5090	5026	4814	4690	6819	6699	6665	6637	6559	6768	6699	6621

	ohhrf	phhrf	qhhrf	rhhrf	shhrf	thhrf	uhhrf	vhhrf	whhrf	xhhrf	yhhrf
p5	1057.20	1025.94	839.60	800.91	519.17	522.56	505.68	512.91	471.48	464.13	459.99
p10	1410.47	1387.50	1123.51	1083.51	696.87	706.24	698.61	705.26	667.67	668.38	667.28
p25	2364.65	2342.26	1757.51	1752.13	1275.65	1285.39	1260.76	1275.52	1281.83	1269.57	1286.98
p50	3954.85	3986.41	2534.35	2742.57	2558.11	2557.68	2523.88	2536.77	2403.33	2458.21	2525.97
p75	6235.95	6529.49	3572.20	4139.08	4169.73	4300.34	4390.05	4500.64	4136.26	4451.72	4731.54
p90	9884.17	10807.73	5126.70	6039.75	6386.98	6753.57	7171.05	7480.33	6858.64	7591.85	8349.94
p95	13169.40	14342.68	6448.46	7832.78	8181.11	9040.54	9880.52	10691.89	9680.28	10574.96	11557.25
N	7492	7220	13091	11796	12320	11909	11644	11294	12361	11552	10921

Kombinierbarkeit der Gewichte

- Querschnittsgewicht in 2008 **Produkt** aus Ziehungs'wkt und Response'wkt in Welle 1 und allen Response'wkt bis 2008.
- Modulares Prinzip
 - 1 Ziehungsdesign → DESIGN
 - 2 Ausfallanalyse Welle 1
 - 3 Post-Stratifizierung Welle 1 → AHHRF
 - 4 Ausfallanalyse Welle 2 → BHBLEIB
 - 5 Post-Stratifizierung Welle 2 → BHHRF
 - 6 ...
 - 7 Ausfallanalyse Welle 25 → YHBLEIB
 - 8 Post-Stratifizierung Welle 25 → YHHRF
- Möglichkeit der individuellen Manipulation der Gewichte
 - Herausrechnen der Design-Informationen: AHHRF/DESIGN
 - Herausrechnen der querschnittl. Post-Stratifizierung: $XHHRF \times YHBLEIB = YHHRFNEU$ statt YHHRF.
 - ...

Typische Anwendungen I: Inferenz bzgl. Querschnitt

- Beispiele
 - Wie hoch ist der Anteil der Personen in den Jahren 1987 bis 1990, die sich große Sorgen um ihren Arbeitsplatz machen?
 - Sind Raucher weniger gesund als Nicht-Raucher? Wie stark ist der Zusammenhang zwischen Rauchen und Gesundheitszustand in den Jahren 2004 und 2006?
- Standardgewichtungsfaktor $\$HHRF$ bzw. $\$PHRF$
 - Überproportionalität v. Teilstichproben berücksichtigt
 - Umfassen jeweils alle Teilstichproben.
 - $\$HHRF$ bzw. $\$PHRF$
- Es gibt Gewichte spezifisch f. Analysen unter Ausschluß der neuen Teilstichproben in Welle 1:
 - A–D ohne E in 1998: $OHHRFAD$ bzw. $OPHRFAD$
 - A–E ohne F in 2000: $QHHRFAE$ bzw. $QPHRFAE$
 - A–F ohne G in 2002: $SHHRFAF$ bzw. $SPHRFAF$
 - A–G ohne H in 2006: $WHHRFAG$ bzw. $WPHRFAG$

Typische Anwendungen II: Inferenz bzgl. Längsschnitt

- Beispiele
 - Wie hoch ist der Anteil der Personen, die sich im Zeitraum 1987 bis 1990 große Sorgen um ihren Arbeitsplatz machen?
 - Gibt es einen Effekt des Rauchens auf die Veränderung des Gesundheitszustands? Verändert sich der Gesundheitszustand von Befragten, die das Rauchen zwischen 2004 und 2006 aufgeben/anfangen?
- Balancierte Panelstichprobe für erstes Bsp.
 - Kehrwert der (geschätzten) Wkt Person in Welle G (und F und E und D) zu beobachten \approx $DPHRF \times EPBLEIB \times FPBLEIB \times GPBLEIB$
- Unbalancierte Panelstichprobe für zweites Bsp.
 - $UPHRF \approx \{1/\widehat{Pr}(c_{i,1} = 1)\}$
 - $UPHRF \times VPBLEIB \approx \{1/\widehat{Pr}(c_{i,1} = 1, c_{i,2} = 1)\}$
 - $UPHRF \times VPBLEIB \times WPBLEIB \approx \{1/\widehat{Pr}(c_{i,1} = 1, c_{i,2} = 1, c_{i,3} = 1)\}$

Gewichtung und Varianzschätzung

- Übliche Schätzverfahren in Statistik-Software unterstellen einfache Zufallsstichproben (d.h., jede Einheit hat gleiche, von Null verschiedene Ziehungs'wkt).
- Designeffekt
 - Vorteile von Klumpung (Anfahrtswege) und Schichtung (hinreichend große Fallzahl für kleine Bev.gruppen) gehen oftmals zu Lasten der Stichprobeneffizienz.
- Effekte der (zusätzlichen) Schätzung der Response'wkt
 - Je besser Vorhersage der Response'wkt, desto höher Varianz der Gewichte.
 - Je höher Varianz der Gewichte, desto geringer Effizienz der Stichprobe.
- Bei der Interpretation sollten Konfidenzintervalle bzw. Varianzen von Schätzwerten berücksichtigt werden.

Designeffekt und Stichprobenfehler

- Idee: Wie hoch ist die Varianz in der Stichprobe, die auf die Gewichte/Klumpung zurückzuführen ist im Verhältnis zur Varianz in den Angaben der Befragten?
- Effektive Stichprobengröße mit je \bar{n}_j SSU's in m PSU's

$$n^* = \frac{m \times \bar{n}_j}{deff}$$

- Designeffekt

$$deff = n \underbrace{\frac{\sum_{i=1}^n (\pi_i^{-1})^2}{(\sum_{i=1}^n \pi_i^{-1})^2}}_{deff_{strat}} \times \underbrace{(1 + (\bar{n}_j - 1) \rho_y)}_{deff_{clust}}$$

- Intraklassen-Korrelation

$$\rho = \frac{\sigma_{PSU}^2}{\sigma_{PSU}^2 + \sigma_{SSU}^2}$$

Beispiel

- Designeffekt durch Schichtung
 - Disproportionale Ziehung von 1000 Ausländern und Deutschen in Ost- und Westdeutschland mit je 250 Befragten pro Schicht.
 - Designeffekt = 1.88
- Designeffekt durch Klumpung
 - Ziehung von je 10 Personen (SSU's) pro 100 regionalen Einheiten (PSU's) führt zur Stichprobe von 1000 Befragungen.
 - Das interessierende Merkmal (z.B. Lebenszufriedenheit) variiert sowohl individuell als auch regional mit einem $\rho = 0.10$.
 - Designeffekt = 1.90
- Effektive Stichprobengröße aufgrund der Schichtung und Klumpung ist (hinsichtlich der Lebenszufriedenheit)
$$n^* = \frac{1000}{1.88 \times 1.90} = 280.$$

Korrektur der Varianzschätzung in STATA

- Berücksichtigung der Schichtung und Stufung

```
svyset PSU , strata(STRAT)
```

- Berücksichtigung der Varianz der Gewichte

```
svyset [weight = WGHT]
```

- Alternative Herangehensweisen bei gestuften Daten
 - Huber-Sandwich-White: `cluster`, `robust`
 - Hierarchische Modelle: `gllamm`, `xtreg`, `xtmixed`

Varianzschätzung nach dem Random Group Konzept

- “Simulation” von (im Falle des SOEP) $R = 8$ unabhängigen, identisch gezogenen (Teil-)Stichproben.
- Berechnung des interessierenden Schätzwertes für jede Teilstichprobe. Streuung dieser R Schätzwerte liefert Basis für Varianzschätzung.
- Vorgehen
 - 1 Berechnung der $R = 8$ Schätzwerte für die Random Groups
 - 2 Anordnung der Einzelergebnisse nach Größe (aufsteigend)
 - 3 Bestimmung Konfidenzintervall:
 - (a) $\Pr(\hat{y}_1 < y < \hat{y}_8) \approx 1 - 0.008$
 - (b) $\Pr(\hat{y}_2 < y < \hat{y}_7) \approx 1 - 0.07$
- Schätzung der Varianz (eine mögliche Variante)

$$\hat{\sigma}_R^2 = \frac{1}{R(R-1)} \sum_{r=1}^R (\hat{y}_r - \bar{\hat{y}})^2 \quad \text{mit} \quad \bar{\hat{y}} = \frac{1}{R} \sum_{r=1}^R \hat{y}_r.$$

Gewichtung in univariaten Analysen?

- 1 Beobachtungs'wkt $\Pr(c_i = 1)$ ist zufällig (MCAR) bzgl. des inhaltlich interessierenden Merkmals y : ungewichtete Analyse führt zu unverzerrter Schätzung von \hat{y} .
- 2 Beobachtungs'wkt $\Pr(c_i = 1)$ ist abhängig von \mathbf{x} (MAR) bzgl. des inhaltlich interessierenden Merkmals y : unter Berücksichtigung von \mathbf{x} durch die Gewichte führt die gewichtete Analyse im Gegensatz zur ungewichteten Analyse zu unverzerrter Schätzung von \hat{y} .

Gewichtung in multivariaten Analysen?

- 3 Beobachtungs'wkt $\Pr(c_i = 1)$ ist abhängig von \mathbf{x} (MAR) und das (Regressions-)Modell von y ist vollständig, d.h. es enthält \mathbf{x} : ungewichtete Analyse führt zu unverzerrter Schätzung von \hat{y} .
- 4 Beobachtungs'wkt $\Pr(c_i = 1)$ ist abhängig von \mathbf{x} (Missing At Random) und das (Regressions-)Modell von y ist unvollständig, d.h. es enthält nicht \mathbf{x} : unter Berücksichtigung von \mathbf{x} durch die Gewichte führt die gewichtete Analyse im Gegensatz zur ungewichteten Analyse zu unverzerrter Schätzung von \hat{y} .
- 5 Beobachtungs'wkt $\Pr(c_i = 1)$ ist abhängig von unbeobachteten Werten von y (NMAR): Modell-basierte Lösungen.

Die Abwägung von Annahmen

- Erste Annahme: Sind Daten MCAR, MAR oder NMAR?
- Zweite Annahme: Ist das Gewichtungsmodell bzgl. der eigenen Frage vollständig?
 - Ziehungs'wkt: ja.
 - Beobachtungs'wkt: hinsichtlich vieler Fragestellung, ja.
- Dritte Annahme: Ist das eigene inhaltliche Modell vollständig?
- Sonderfall MAR
 - Je relevanter Gewichtungsmodell für eigene Frage und je unvollständiger inhaltliches Modell, desto wichtiger wird die Gewichtung für die Verhinderung von Verzerrungen.
 - Je weniger relevant Gewichtungsmodell für eigene Frage und je vollständiger inhaltliches Modell, desto weniger hilfreich ist die Gewichtung für die Verhinderung von Verzerrungen bei gleichzeitiger Ineffizienz.
- **Gewichtung ist die konservativere Strategie und somit vorzuziehen.**

Literatur: Allgemein

- Kalton, G. & Brick, J.M. (1995). Weighting Schemes for Household Panel Surveys. *Survey Methodology*, 21, 33–44.
- Wolter, K.M. (1995). *Introduction to Variance Estimation*. Springer: New York.
- Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Literatur: Hochrechnung SOEP

Auf <http://www.diw.de/deutsch/sop/service/doku/index.html>

Haisken-DeNew, J.P. & Frick, J.R. (2005). DTC — Desktop Companion to the German Socio-Economic Panel (SOEP).

Pischner, R. (2007) Die Querschnittsgewichtung und die Hochrechnungsfaktoren des Sozio-oekonomischen Panels (SOEP) ab Release 2007 (Welle W): Modifikationen und Aktualisierungen. DIW Data Documentation 22.

Pischner, R. (2002). Die Hochrechnung der ersten Welle der Stichprobe F des SOEP.

Pischner, R. (2003). Integrated Cross-sectional Weighting (Sub-samples A-G) for 2003.

Literatur: Hochrechnung SOEP

- Kroh, M. (2009). Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (GSOEP) (1984-2008). DIW Data Documentation 47. Berlin: DIW.
- Spiess, M. & Rendtel, U. (2000). Combining an ongoing panel with a new cross-sectional sample, DIW Discussion Paper No. 198.
Webpage-Titel: Martin Spieß – Sample E Integration into SOEP.
- Spiess, M. (2001). Combining the SOEP ongoing panel (subsamples A-E) with the new subsample F. Webpage-Titel: Sample F Integration into SOEP.
- Spiess, M. (2000). Derivation of design weights: The case of the German Socio-Economic Panel (GSOEP). DIW Discussion Paper No. 197.

Literatur: Hochrechnung SOEP

Sonstige SOEP-spezifische Literatur

- Rendtel, U. (1995). Panelmortalität und Panelrepräsentativität. Frankfurt: Campus.
- Rendtel, U., Pannenberg, M. & Daschke, S. (1997). Die Gewichtung der Zuwanderer-Stichprobe des Sozio-ökonomischen Panels (SOEP). *Vierteljahrsheft für Wirtschaftsforschung* 2/97, 271–286.
- Rendtel, U., Wagner, G. & Frick, J. (1995). Eine Strategie zur Kontrolle von Längsschnittgewichten in Panelerhebungen - Das Beispiel des Sozio-ökonomischen Panels (SOEP). *Allgemeines Statistisches Archiv*, 79, 252–277.